Int. J. Cancer: **116,** 768–773 (2005)

Polymorphisms in the DNA nucleotide excision repair genes and lung cancer risk in Xuan Wei, China

Min Shen[1]*, Sonja I. Berndt[1], Nathaniel Rothman[1], David M. DeMarini[2], Judy L. Mumford[2], Xingzhou He[3], Matthew R. Bonner[1], Linwei Tian[4], Meredith Yeager[1], Robert Welch[1], Stephen Chanock[1], Tongzhang Zheng[5], Neil Caporaso[1] and Qing Lan[1]

[1]*Division of Cancer Epidemiology and Genetics, National Cancer Institute, National Institutes of Health, Bethesda, MD, USA*
[2]*National Health and Environmental Effects Research Laboratory, U.S. Environmental Protection Agency, Research Triangle Park, NC, USA*
[3]*Chinese Center for Disease Control and Prevention, Beijing, People's Republic of China*
[4]*School of Public Health, University of California, Berkeley, CA, USA*
[5]*Yale School of Public Health, Yale University, New Haven, CT, USA*

The lung cancer mortality rate in Xuan Wei County is among the highest in China and has been attributed to exposure to indoor smoky coal emissions that contain very high levels of polycyclic aromatic hydrocarbons (PAHs). Nucleotide excision repair (NER) plays a key role in reversing DNA damage from exposure to environmental carcinogens, such as PAHs, that form bulky DNA adducts. We studied single nucleotide polymorphisms (SNPs) and their corresponding haplotypes in 6 genes (*ERCC1, ERCC2/XPD, ERCC4/XPF, ERCC5/XPG, RAD23B* and *XPC*) involved in NER in a population-based case-control study of lung cancer in Xuan Wei. A total of 122 incident primary lung cancer cases and 122 individually matched controls were enrolled. Three linked SNPs in *ERCC2* were associated with lung cancer with similar ORs; *e.g.*, persons with the Gln allele at codon 751 had a 60% reduction of lung cancer (OR = 0.40, 95% CI 0.18–0.89). Moreover, one haplotype in *ERCC2* was associated with a decreased risk of lung cancer (OR = 0.40, 95% CI 0.19–0.85) compared to the most common haplotype. In addition, subjects with one or 2 copies of the Val allele at codon 249 of *RAD23B* had a 2-fold increased risk of lung cancer (OR = 1.91, 95% CI 1.12–3.24). In summary, our results suggest that genetic variants in genes involved in the NER pathway may play a role in lung cancer susceptibility in Xuan Wei. However, due to the small sample size, additional studies are needed to evaluate these associations within Xuan Wei and in other populations with substantial environmental exposure to PAHs.

Key words: lung cancer; DNA repair; single nucleotide polymorphism; nucleotide excision repair; polycyclic aromatic hydrocarbon

Lung cancer is the leading cause of death from cancer worldwide, with an estimated mortality in 2000 of 31.4 per 100,000 for men and 9.5 per 100,000 for women.[1] Overall, lung cancer incidence and mortality rates are higher for males than females, due largely to the higher prevalence of tobacco smoking among men. Compared to worldwide patterns and those in China, lung cancer has distinctive characteristics in rural Xuan Wei County, Yunnan Province, China. Lung cancer mortality in Xuan Wei has been reported to be 8 times the Chinese national average for women and 4 times that for men.[2] Although very few women smoke, the lung cancer mortality rates in Xuan Wei County were similar between men and women (27.7 and 25.3 per 100,000 for males and females, respectively).[2] This pattern has been attributed to burning smoky coal indoors for heating and cooking without adequate ventilation with high exposure time, particularly among women, accounting for >90% of lung cancer cases for both men and women.[2,3] When smoky coal is burned, the indoor air concentration of particulate matter and extractable organic matter may be as high as 24.4 and 17.6 mg/m^3,[2] respectively, and the corresponding benzo[*a*]pyrene concentration, an indicator of carcinogenic PAHs, can reach as high as 19.3 μg/m^3,[4] which is comparable to exposure levels experienced by coke oven workers. As such, the study of lung cancer in Xuan Wei provides a unique opportunity to evaluate genetic susceptibility in a non-smoking model of PAH carcinogenesis.

The removal and repair of DNA damage plays a key role in protecting the integrity of the genome from the insults of cancer-causing agents. Several different DNA repair pathways exist, including BER, NER, double strand break repair and mismatch repair.[5] The NER pathway repairs bulky DNA adducts induced by chemical carcinogens, such as PAHs found in smoky coal emissions and tobacco smoke.[6] Several studies have assessed the relationship between SNPs in NER genes and the risk of lung cancer, but the results have been inconsistent.[7–10] Differences in the interaction between environmental exposures and genetic risk factors could contribute to some of the heterogeneity between studies, and there is some evidence that effects may vary as a result of different levels and possibly different types of exposure to carcinogens.[11,12] As exposure to smoky coal emissions is the primary risk factor for lung cancer in Xuan Wei,[2] we hypothesized that genetic variation in the NER pathway would play an important role in lung carcinogenesis in this special population. Here, we report the direction and magnitude of associations between genetic variants in 6 NER genes (*ERCC1, ERCC2/XPD, ERCC4/XPF, ERCC5/XPG, RAD23B* and *XPC*) and lung cancer risk in a population-based case-control study in Xuan Wei, China.

Material and methods

Study population

This was a population-based case-control study of lung cancer in Xuan Wei, China. Details of the study have been described elsewhere.[13] Briefly, the field phase of the study lasted from March 1995 to March 1996. A total of 122 newly diagnosed lung cancer

Abbreviations: BER, base excision repair; CI, confidence interval; EM, estimation-maximization; ERCC1, excision repair cross-complementation group 1; EX, exon; HWE, Hardy-Weinberg equilibrium; IVS, intervening sequence; LD, linkage disequilibrium; NER, nucleotide excision repair; OR, odds ratio; PAH, polycyclic aromati hydrocarbon; RAD23B, RAD23 homolog B; SNP, single nucleotide polymorphism; TFIIH, transcription factor IIH; XPC, xeroderma pigmentosum, complementation group C.

The research described here has been reviewed by the National Environmental and Health Effects Research Laboratory of the U.S. Environmental Protection Agency and approved for publication. Approval does not signify that the contents necessarily reflect the views and policies of the agency nor does mention of trade names or commercial products constitute endorsement or recommendation for use.

***Correspondence to:** Occupational and Environmental Epidemiology Branch, Division of Cancer Epidemiology and Genetics, National Cancer Institute, NIH, DHHS, MSC 7240, 6120 Executive Blvd., Bethesda, MD 20892-7240, USA. Fax: +301-402-1819. E-mail: shenmi@mail.nih.gov

Received 28 September 2004; Accepted after revision 2 February 2005
DOI 10.1002/ijc.21117
Published online 22 April 2005 in Wiley InterScience (www.interscience.wiley.com).

cases were recruited, and the criteria for inclusion were positive histology or cytology results (105 cases) or clinically diagnosed cases who died within the 1-year period (17 cases). Within 2 weeks after the diagnosis and recruitment of each lung cancer case, a matched control was selected randomly from a list of registered households in the same villages. Participation rates for cases and controls were 98% and 100%, respectively. Matching conditions included sex, age (±2 years) and type of fuel currently used for cooking and home heating. A standardized structured questionnaire was used to obtain information about demographic characteristics, lifetime use of different types of coal, tobacco smoking, family history of lung cancer and personal medical history. The study was conducted according to the recommendations for human subject protection of the World Medical Association Declaration of Helsinki. The research protocol was approved by a U.S. Environmental Protection Agency Human Subjects Research Review official for international research projects, and informed consent was obtained from all subjects.

TABLE I – DISTRIBUTION OF DEMOGRAPHIC FEATURES IN LUNG CANCER CASES AND CONTROLS[3]

	Cases (%) (n = 122)	Controls (%) (n = 122)	p value[1]
Age (years)			
<55	52 (43)	51 (42)	
≥55	70 (57)	71 (58)	0.90
Sex			
Male	79 (65)	79 (65)	
Female	43 (35)	43 (35)	1.00
Pack-years[2]			
Nonsmokers	9 (11)	10 (13)	
<25	27 (34)	36 (46)	
≥25	43 (54)	33 (42)	0.26
Smoky coal use (tons)			
<130	51 (42)	72 (59)	
≥130	71 (58)	50 (41)	0.007

[1]Two-sided p value based on χ^2 test.–[2]Males only.–[3]Demographic data were previously reported.

Genotyping

Sputum samples were collected from cases and controls. DNA from sputum samples was extracted using phenol-chloroform extraction,[14] and 15 SNPs in 6 NER genes were genotyped by real-time PCR on an Applied Biosystems (Foster City, CA) 7900HT sequence detection system at the Core Genotyping Facility of the National Cancer Institute as described on the SNP500 website (http://snp500cancer.nci.nih.gov).[15] Of the 122 cases and 122 controls, DNA was successfully extracted from 119 cases and 113 controls, and >95% of DNA samples were successfully genotyped for all candidate SNPs, except *ERCC5* Leu700Leu (88%) and *XPC* EX 16 +315 C>G (93%). Concordance rates between quality-control samples were 99–100% for all assays.

Statistical analysis

An ever-smoker was defined as a smoker of at least one cigarette per day for 6 months or longer. Cut-off points of distribution of smoky coal use (tons) and tobacco smoking (pack-years) were estimated based on the distribution of lifetime cumulative use in controls. The HWE for each SNP was tested using a Pearson χ^2 test or an exact test (if one of the genotypes was infrequent) in controls. Measures of pairwise LD between SNPs within the same gene were carried out with HaploView (http://www.broad.mit.edu/personal/jcbarret/haploview/). Genotype data were analyzed using homozygotes for common allele as the reference group. As genotype data were not obtained for all cases and controls, unconditional logistic regression was used to estimate the ORs and 95% CIs, with 2-sided p values for the association between lung cancer risk and the SNPs adjusted for age, sex and current fuel type, and in an alternative model, pack-years of smoking (≥25 vs. <25) and smoky coal use (≥130 tons vs. <130 tons). The haplotype block structure for each gene with more than one SNP was examined with HaploView using the 4-gamete rule[16] with a minimum frequency of 0.005 for the fourth gamete. For each haplotype block, individual haplotypes were estimated using the EM algorithm, and overall differences in haplotype frequencies between cases and controls were assessed using the omnibus test in SAS/Genetics (SAS Institute, Cary, NC). The association between each haplotype and lung cancer risk was estimated using unconditional logistic regression, with the most common haplotype or the haplotype containing the common alleles as the reference. Data were analyzed with Statistical Analysis Software, version 8.02 (SAS Institute), if not specified elsewhere.

Results

Demographic features, including age, sex, ethnicity, education level, household income, dwelling type and type of fuel source, were comparable between cases and controls (Table I). About 93% of men smoked tobacco, while only one woman smoked. Compared with studies in other populations, the impact of tobacco smoking in Xuan Wei was quite weak, with a 1.7-fold (95% CI 0.8–3.5) risk of lung cancer for exposure to more than 25 pack-years among men, which is consistent with previous studies in Xuan Wei.[17] However, smoky coal use was a strong risk factor in Xuan Wei. Compared to subjects who used <130 tons of smoky coal during their lifetime, heavy smoky coal users (≥130 tons) had a 2.27-fold (95% CI 1.25–4.10) increased risk of lung cancer.

Fifteen SNPs were genotyped in 6 NER genes. These SNPs were comprised of nonsynonymous and synonymous substitutions in coding regions, as well as single base pair changes in noncoding regions. With the exception of *ERCC5* His46His (p = 0.04) and *XPC* Lys939Gln (p = 0.03), the genotype frequencies for all of the SNPs in controls were consistent with HWE. Quality-control samples for the 2 SNPs out of HWE were rechecked and the concordance rate was >99% for each. Genotype frequencies for cases and controls and the main effect of these SNPs on lung cancer risk are shown in Table II. Five SNPs in 3 genes displayed a significant or borderline significant association with lung cancer. Three linked SNPs in *ERCC2* were associated with decreased risk of lung cancer. Homozygotes of a minor allele of *ERCC2* Lys751Gln and *ERCC2* IVS19 –70 C>T were rare, and individuals with at least one variant had a 60% reduction of lung cancer risk. The *ERCC2* Arg156Arg polymorphism showed a borderline association with lung cancer, with the *CC* genotype displaying a lower risk of lung cancer compared to the *AA* genotype. Subjects with at least one 249Val allele for *RAD23B* had an approximately 2-fold higher risk of lung cancer, and *XPC* 939Gln homozygotes displayed an increased risk of lung cancer that was borderline significant (OR = 2.45, 95% CI 0.96–6.21, p = 0.06). Because *RAD23B* and *XPC* work collectively to recognize DNA damage as a primary step in NER,[6] we analyzed the combined effects of the *RAD23B* Ala249Val and *XPC* Lys939Gln polymorphisms and found that subjects with both the *XPC* 939Gln/Gln genotype and either the *RAD23B* 249Ala/Val or 249Val/Val genotype had a 6-fold higher risk of lung cancer (Table III). Other variants evaluated in this study were not associated with lung cancer risk.

We examined pairwise LD between the SNPs and the haplotype block structure for genes in which more than one SNP was genotyped. The extent of LD varied between the SNPs for *ERCC2, ERCC5* and *XPC* (Table IV); and for both *ERCC2* and *ERCC5,* we observed 2 different haplotype blocks within each gene. For *ERCC2*, Lys751Gln and IVS19 –70 C>T appeared to be in one haplotype block using the 4-gamete rule, whereas Asp312Asn and Arg156Arg were located in a second block. Similarly, the His46His and Cys529Ser variants in *ERCC5* were in a separate haplotype block from Leu700Leu and His1104Asp. For *XPC,* only Lys939Gln and Ala499Val were in the same haplotype block. The SNPs analyzed in *ERCC1* were not tightly linked (D' = 0.77) and did not appear to be in the same haplotype block.

Haplotypes were estimated for SNPs within the same haplotype block. The distribution of estimated haplotypes and results of

TABLE II – MAIN EFFECT OF SNPs OF DNA NER GENES ON LUNG CANCER RISK IN XUAN WEI, CHINA

Gene and SNP position (db SNP ID)	Cases (%) (n = 119)	Controls (%) (n = 113)	OR[1]	95% CI	p value[3]	OR[2]	95% CI	p value[3]
ERCC1								
IVS5 +33 C>A (rs3212961)								
CC	41 (35)	34 (32)	Ref.			Ref.		
CA	58 (49)	54 (50)	0.90	0.50–1.61	0.71	0.87	0.48–1.59	0.65
AA	19 (16)	19 (18)	0.83	0.38–1.81	0.63	0.82	0.37–1.83	0.63
CA+AA	77 (65)	73 (68)	0.88	0.50–1.53	0.65	0.86	0.48–1.51	0.59
Trend					0.61			0.59
IVS3 +74 G>C (rs3212948)								
GG	66 (56)	67 (60)	Ref.			Ref.		
GC	45 (38)	40 (36)	1.15	0.66–2.00	0.62	1.13	0.64–1.99	0.67
CC	7 (6)	5 (4)	1.44	0.43–4.80	0.55	1.66	0.49–5.62	0.42
GC+CC	52 (44)	45 (40)	1.18	0.70–2.01	0.53	1.19	0.69–2.04	0.54
Trend					0.48			0.43
ERCC2								
Ex23 +61 A>C (rs1052559)								
Lys751Gln								
AA (Lys/Lys)	107 (91)	86 (80)	Ref.			Ref.		
AC (Lys/Gln)	11 (9)	20 (19)	0.45	0.20–0.99	0.048	0.42	0.18–0.95	0.037
CC (Gln/Gln)		2 (2)						
AC (Lys/Gln) + CC (Gln/Gln)	11 (9)	22 (20)	0.40	0.18–0.89	0.02	0.39	0.17–0.86	0.02
IVS19 −70 C>T (rs1799787)								
CC	106 (91)	89 (80)	Ref.			Ref.		
CT	11 (9)	20 (18)	0.46	0.21–1.03	0.059	0.44	0.19–0.99	0.047
TT		2 (2)						
CT+TT	11 (9)	22 (20)	0.42	0.19–0.92	0.03	0.40	0.18–0.90	0.03
Ex10 −16 G>A (rs1799793)								
Asp312Asn								
GG (Asp/Asp)	109 (92)	99 (88)	Ref.			Ref.		
GA (Asp/Asn)	9 (8)	14 (12)	0.58	0.24–1.40	0.23	0.62	0.25–1.53	0.30
Ex6 −10 A>C (rs238406)								
Arg156Arg								
AA	30 (26)	24 (22)	Ref.			Ref.		
AC	64 (55)	50 (45)	1.01	0.52–1.96	0.98	1.02	0.52–2.01	0.96
CC	23 (20)	37 (33)	0.50	0.23–1.05	0.06	0.47	0.22–1.01	0.05
AC+CC	87 (74)	87 (78)	0.79	0.42–1.46	0.45	0.77	0.41–1.46	0.42
Trend					0.058			0.046
ERCC4								
Ex11 −247 T>C (rs1799801)								
Ser835Ser								
TT	69 (59)	74 (67)	Ref.			Ref.		
TC	40 (34)	32 (29)	1.35	0.76–2.42	0.30	1.33	0.73–2.40	0.35
CC	8 (7)	5 (5)	1.70	0.53–5.48	0.37	1.71	0.52–5.58	0.38
TC+CC	48 (41)	37 (33)	1.40	0.81–2.43	0.22	1.38	0.79–2.41	0.26
Trend					0.21			0.23
ERCC5								
Ex2 +50 T>C (rs1047768)								
His46His								
TT	55 (47)	63 (56)	Ref.			Ref.		
TC	49 (42)	36 (32)	1.56	0.89–2.75	0.12	1.54	0.87–2.74	0.14
CC	14 (12)	13 (12)	1.23	0.53–2.85	0.63	1.31	0.55–3.07	0.54
TC+CC	63 (53)	49 (44)	1.47	0.88–2.48	0.14	1.48	0.87–2.52	0.15
Trend					0.28			0.25
Ex8 −369 G>C (rs2227869)								
Cys529Ser								
GG (Cys/Cys)	103 (87)	100 (90)	Ref.			Ref.		
GC (Cys/Ser)	14 (12)	11 (10)	1.25	0.54–2.91	0.60	1.34	0.57–3.19	0.50
CC (Ser/Ser)	1 (1)							
Ex9 −100 C>A (rs2228959)								
Leu700Leu								
CC	97 (92)	88 (89)	Ref.			Ref.		
CA	9 (8)	11 (11)	0.75	0.29–1.91	0.54	0.79	0.30–2.06	0.63
Ex15 −344 C>G (rs17655)								
His1104Asp								
CC (His/His)	38 (33)	38 (35)	Ref.			Ref.		
CG (His/Asp)	52 (45)	46 (42)	1.16	0.63–2.12	0.63	1.08	0.58–2.01	0.81
GG (Asp/Asp)	26 (22)	25 (23)	1.05	0.52–2.14	0.89	1.03	0.49–2.13	0.95
CG (His/Asp) + GG (Asp/Asp)	78 (67)	71 (65)	1.12	0.64–1.95	0.69	1.06	0.60–1.88	0.84
Trend					0.85			0.92
RAD23B								
Ex7 +65 C>T (rs1805329)								
Ala249Val								
CC (Ala/Ala)	58 (49)	72 (65)	Ref.			Ref.		
CT (Ala/Val)	52 (44)	34 (31)	1.90	1.09–3.30	0.02	1.79	1.02–3.16	0.04
TT (Val/Val)	8 (7)	5 (5)	1.98	0.61–6.41	0.26	1.88	0.56–6.25	0.30
CT (Ala/Val)+ TT (Val/Val)	60 (51)	39 (35)	1.91	1.12–3.24	0.02	1.81	1.05–3.11	0.03
Trend					0.03			0.05

TABLE II – MAIN EFFECT OF SNPs OF DNA NER GENES ON LUNG CANCER RISK IN XUAN WEI, CHINA (CONTINUED)

Gene and SNP position (db SNP ID)	Cases (%) (n = 119)	Controls (%) (n = 113)	OR[1]	95% CI	p value[3]	OR[2]	95% CI	p value[3]
XPC								
Ex16 +315 C>G (rs2229090)								
CC	52 (48)	47 (44)	Ref.			Ref.		
CG	52 (48)	52 (49)	0.91	0.52–1.58	0.74	0.90	0.51–1.58	0.70
GG	5 (5)	8 (7)	0.58	0.18–1.93	0.38	0.56	0.16–1.91	0.35
CG+GG	57 (52)	60 (56)	0.87	0.51–1.49	0.61	0.85	0.49–1.48	0.57
Trend					0.45			0.42
Ex16 +211 A>C (rs2228001)								
Lys939Gln								
AA (Lys/Lys)	43 (38)	39 (37)	Ref.			Ref.		
AC (Lys/Gln)	50 (44)	58 (55)	0.78	0.44–1.38	0.39	0.67	0.37–1.23	0.20
CC (Gln/Gln)	21 (18)	8 (8)	2.45	0.96–6.21	0.06	2.22	0.86–5.74	0.10
AC (Lys/Gln) + CC (Gln/Gln)	71 (62)	66 (63)	0.98	0.56–1.69	0.93	0.86	0.48–1.52	0.60
Trend					0.25			0.43
Ex9 −377 C>T (rs2228000)								
Ala499Val								
CC (Ala/Ala)	56 (48)	50 (45)	Ref.			Ref.		
CT (Ala/Val)	47 (41)	47 (43)	0.89	0.51–1.56	0.69	0.91	0.52–1.62	0.75
TT (Val/Val)	13 (11)	13 (12)	0.89	0.38–2.11	0.79	0.98	0.41–2.38	0.97
CT (Ala/Val) + TT (Val/Val)	60 (52)	60 (55)	0.89	0.53–1.51	0.67	0.93	0.54–1.59	0.78
Trend					0.70			0.86

[1]Adjusted for age, sex and current fuel type by unconditional logistic regression.–[2]Adjusted for age, sex, current fuel type, pack-years of smoking and smoky coal use by unconditional logistic regression.–[3]Two-sided p value.

TABLE III – COMBINED GENOTYPES OF *RAD23B* ALA249VAL AND *XPC* LYS939GLN ON LUNG CANCER RISK

RAD23B Ex7 +65 C>T (rs1805329) Ala249Val	*XPC* Ex16 +211 A>C (rs2228001) Lys939Gln	Cases (n = 113)	Controls (n = 103)	OR	95% CI	p value[1]
CC	AA/AC	45 (37)	60 (49)	Ref.		
CC	CC	58 (47)	41 (34)	1.91	1.09–3.34	0.02
CT/TT	AA/AC					
CT/TT	CC	10 (8)	2 (2)	6.35[2]	1.26–62.54	0.02
						0.002[3]

[1]Two-sided p value. The test for multiplicative interaction between these 2 variants was not significant.–[2]By exact logistic regression.–[3]Trend test.

TABLE IV – LEWONTIN'S D′, A MEASURE OF LD BETWEEN SNPs IN *ERCC2, ERCC5*, AND *XPC*

	ERCC2			*ERCC5*			*XPC*	
	Lys751Gln	IVS18−70 C>T	Asp312Asn	His46His	Cys529Ser	Leu700Leu	Ex16 +315 C>G	Lys939Gln
ERCC2								
IVS19−70 C>T	1.0							
Asp312Asn	0.57	0.57						
Arg156Arg	0.60	0.61	1.0					
ERCC5								
Cys529Ser				1.0				
Leu700Leu				0.19	0.08			
His1104Asp				0.51	1.0	1.0		
XPC								
Lys939Gln							0.71	
Ala499Val							1.0	1.0

logistic regression are shown in Table V. An omnibus likelihood ratio test showed a significant difference in frequency profiles of an *ERCC2* haplotype (Lys751Gln, IVS19 –70 C>T) between cases and controls (p = 0.01). However, the Lys751Gln and IVS19 –70 C>T SNPs in *ERCC2* were completely linked, so the estimated OR for the joint association of the variants was the same as for the individual SNPs (OR = 0.40, 95% CI 0.19–0.85). The omnibus test for the haplotype block containing *ERCC2* Asp312Asn and Arg156Arg was not statistically significant (p = 0.11).

Discussion

The high mortality of lung cancer in Xuan Wei, China, is overwhelmingly driven by exposure to PAH-rich coal combustion emissions.[2,3] Further, it has been shown that DNA base pair changes in *TP53* and *KRAS* from lung tumors of nonsmokers exposed to PAH-rich coal emissions in Xuan Wei have a mutational spectrum highly consistent with PAH mutagenesis, which is distinctive from the mutation pattern in lung tumors caused by cigarette smoke.[18] Here, we report that genotypes of 2 NER genes (*ERCC2* and *RAD23B*), which play a key role in repairing DNA damaged by PAHs, altered risk of lung cancer in this population. In addition, the *XPC* Gln939Gln genotype was associated with a borderline significant risk of lung cancer.

The protein encoded by *ERCC2* is involved in transcription-coupled NER and is an integral member of the basal transcription factor TFIIH complex,[19] which is necessary for normal transcription initiation of NER. The 751Gln variant of the *ERCC2* gene leads to a conformational change in the coded protein at the

TABLE V – HAPLOTYPE ANALYSIS OF DNA NER GENES ON LUNG CANCER RISK IN XUAN WEI, CHINA

	Haplotypes	Cases	Controls	OR[1]	95% CI	*p* value[2]
ERCC2	Lys751Gln–IVS18-70 C>T					
Hap1	A-C	223	192	Ref.		
Hap2	C-T	11	24	0.40	0.19–0.85	0.02
Omnibus test						0.01
ERCC2	Asp312Asn–Arg156Arg					
Hap1	G-A	125	100	Ref.		
Hap2	G-C	102	112	0.73	0.50–1.06	0.95
Hap3	A-C	9	14	0.70	0.29–1.70	0.24
Omnibus test						0.11
ERCC5	His46His–Cys529Ser					
Hap1	T-G	159	164	Ref.		
Hap2	C-G	61	51	1.24	0.80–1.90	0.96
Hap3	C-C	16	11	1.48	0.66–3.32	0.48
Omnibus test						0.47
ERCC5	Leu700Leu–His1104Asp					
Hap1	C-C	130	126	Ref.		
Hap2	C-G	97	89	1.06	0.73–1.55	0.53
Hap3	A-G	9	11	0.80	0.32–1.99	0.58
Omnibus test						0.75
XPC	Lys939Gln–Ala499Val					
Hap1	C-C	95	80	Ref.		
Hap2	A-C	67	69	0.82	0.52–1.28	0.62
Hap3	A-T	74	75	0.83	0.53–1.28	0.65
Omnibus test						0.44

[1]Adjusted for age, sex and current fuel type by unconditional logistic regression.–[2]Two-sided *p* value.

domain of interaction between the ERCC2 protein and its helicase activator, p44 protein, inside the TFIIH complex.[20] The role of the *ERCC2* polymorphism in human cancer is unclear, with equivocal results from both functional and molecular epidemiologic studies.[11,21–31] The 751Gln allele has been associated with higher DNA adduct levels, lowered DNA repair capacity or increased chromosomal aberrations in some studies.[21,24,25] However, the association between higher DNA adduct levels and the 751Gln allele was restricted to specific exposure subgroups, *e.g.,* never-smokers[23] and traffic-exposed workers[26] in 2 studies. Other studies found null[27,28] or even protective[22,29] effects for this variant.

Similarly, epidemiologic studies have not yielded a clear picture of the effect of the *ERCC2* 751Gln allele on lung cancer risk. A meta-analysis from 9 published case-control studies showed a small increased risk (OR = 1.21, 95% CI 1.02–1.43) associated with 751Gln.[10] However, there was no significant association among Asians, and there was evidence of significant heterogeneity among the 3 studies, indicating uncertainty of the role of this SNP in Asian populations. Differences in associations between Caucasian and Asian populations could be due to differences in LD patterns between the 2 populations if the *ERCC2* 751Gln allele is merely linked to the true ''at-risk'' variant. A study of *ERCC2* haplotypes in 3 different ethnic groups revealed that the haplotype structure of *ERCC2* differed substantially between Europeans, Asians and Africans.[32]

Alternatively, differences in exposures may alter the effect of the *ERCC2* polymorphism on lung cancer risk. One study in Caucasians by Zhou et al.[11] found that the risk associated with the *ERCC2* 751Gln allele decreased as pack-years increased and the 751Gln allele had a protective effect among heavy smokers, opposite to its effect among nonsmokers. In contrast, a case-control study in China found that the association between the 751Gln allele and lung cancer was greater among heavy smokers.[33] However, the minimum number of pack-years for heavy smokers was much lower in the Chinese study compared to the Caucasian study ($\geq$29 pack-years *vs.* >55 pack-years), and in the Caucasian study, persons with the 751Gln allele who smoked 25–55 pack-years also displayed an increased risk of lung cancer.[11] Thus, it is likely that the protective effects of the 751Gln allele are seen only at very high exposure levels. We found that the *ERCC2* 751Gln allele as well as 2 other linked variants in *ERCC2* were associated with a reduced risk of lung cancer. In view of the overall heavy exposure to PAHs in Xuan Wei, our finding is consistent with the results of Zhou *et al.*[11] and provides a novel lead for further study on the role of the *ERCC2* Lys751Gln polymorphism.

The protein encoded by *RAD23B* is one of 2 human homologs of *Saccharomyces cerevisiae Rad23*. RAD23B and XPC bind to form an XPC-HHR23 heterodimeric subcomplex, which plays a key part in DNA damage recognition in the NER global genome repair pathway.[6] We found that the *RAD23B* Val allele was associated with increased risk of lung cancer; however, its biologic function is not clear, and additional studies evaluating the functional significance of this polymorphism are warranted.

XPC encodes a 940–amino acid protein that stably combines with RAD23B in the DNA damage recognition step of NER. Laboratory studies show that $XPC^{-/-}$ mice have an increased risk of chemically induced lung tumors compared to normal or heterozygous mice.[34] Even though an *XPC 939* allele-specific complementation assay utilizing post-UV host cell reactivation did not find different DNA repair capacity between the 2 alleles,[35] carriers with homozygotes of Gln/Gln at codon 939 were found to have a 2.5-fold increased risk ($p = 0.06$) of lung cancer compared to those with Lys/Lys and Gln/Lys genotypes in our study. Although this finding could be a false-positive, the association may be due to LD with another polymorphism at intron 11 that is located at a splice acceptor site and associated with decreased DNA repair activity.[36] Other studies have examined a biallelic poly(AT) insertion/deletion polymorphism *(XPC-PAT)* which is in LD with the functional splice site polymorphism at intron 11. Although one study in a Chinese population did not observe an association with the *XPC-PAT* polymorphism,[37] a study in Caucasians found an increased risk of lung cancer for the $PAT^{+/+}$ genotype.[38] These findings suggest that variants in *XPC* may alter the risk of lung cancer. Since RAD23B and XPC unite in the DNA damage recognition step of NER, variants in both genes may interact to hinder NER and increase the risk of lung cancer. Our study found that persons with ''at-risk'' genotypes for both *RAD23B* and *XPC* had a significantly higher risk of lung cancer, which is consistent with this hypothesis.

Our study is limited by its small sample size and, consequently, low power to detect effects that may truly exist. Also, given the borderline significance of some associations and multiple comparisons carried out, there is a possibility that one or more findings

are false-positives.[39] As such, our results need to be considered as preliminary. However, these findings are biologically plausible and derive from a unique population that provides a special model of nonsmoking PAH carcinogenesis. Also, in addition to providing further elucidation of the role of *ERCC2* in lung cancer, this study points to 2 new variants (*XPC* Lys939Gln and *RAD23B* Ala249Val) in the NER pathway that deserve further exploration and research in other studies.

In summary, we found that genetic polymorphisms in *ERCC2, RAD23B* and *XPC* were associated with lung cancer risk in Xuan Wei, China. This suggests that NER may play a role in the pathogenesis of lung cancer, particularly in populations exposed to high levels of PAHs. A substantially larger case-control study of lung cancer will begin later this year in this region of China and will provide an opportunity to replicate and extend these findings.

References

1. Ferlay J, Bray F, Pisani P, Parkin DM. GLOBOCAN 2000. Cancer incidence, mortality and prevalence worldwide. Lyon: IARC, 2001.
2. Mumford JL, He XZ, Chapman RS, Cao SR, Harris DB, Li XM, Xian YL, Jiang WZ, Xu CW, Chuang JC, Wilson WE, Cooke M. Lung cancer and indoor air pollution in Xuan Wei, China. Science 1987;235: 217–20.
3. He X, Yang R. Lung cancer and indoor air pollution from coal burning. Kun Ming: Yunnan Science and Technology Press, 1994.
4. Mumford JL, Helmes CT, Lee XM, Seidenberg J, Nesnow S. Mouse skin tumorigenicity studies of indoor coal and wood combustion emissions from homes of residents in Xuan Wei, China with high lung cancer mortality. Carcinogenesis 1990;11:397–403.
5. Wood RD, Mitchell M, Sgouros J, Lindahl T. Human DNA repair genes. Science 2001;291:1284–9.
6. Friedberg EC. How nucleotide excision repair protects against cancer. Nature Rev Cancer 2001;1:22–33.
7. Kiyohara C, Otsu A, Shirakawa T, Fukuda S, Hopkin JM. Genetic polymorphisms and lung cancer susceptibility: a review. Lung Cancer 2002;37:241–56.
8. Goode EL, Ulrich CM, Potter JD. Polymorphisms in DNA repair genes and associations with cancer risk. Cancer Epidemiol Biomarkers Prev 2002;11:1513–30.
9. Benhamou S, Sarasin A. *ERCC2/XPD* gene polymorphisms and cancer risk. Mutagenesis 2002;17:463–9.
10. Hu Z, Wei Q, Wang X, Shen H. DNA repair gene *XPD* polymorphism and lung cancer risk: a meta-analysis. Lung Cancer 2004;46:1–10.
11. Zhou W, Liu G, Miller DP, Thurston SW, Xu LL, Wain JC, Lynch TJ, Su L, Christiani DC. Gene–environment interaction for the *ERCC2* polymorphisms and cumulative cigarette smoking exposure in lung cancer. Cancer Res 2002;62:1377–81.
12. Zhou W, Liu G, Miller DP, Thurston SW, Xu LL, Wain JC, Lynch TJ, Su L, Christiani DC. Polymorphisms in the DNA repair genes *XRCC1* and *ERCC2,* smoking, and lung cancer risk. Cancer Epidemiol Biomarkers Prev 2003;12:359–65.
13. Lan Q, He X, Costa DJ, Tian L, Rothman N, Hu G, Mumford JL. Indoor coal combustion emissions, *GSTM1* and *GSTT1* genotypes, and lung cancer risk: a case-control study in Xuan Wei, China. Cancer Epidemiol Biomarkers Prev 2000;9:605–8.
14. Garcia-Closas M, Egan KM, Abruzzo J, Newcomb PA, Titus-Ernstoff L, Franklin T, Bender PK, Beck JC, Le Marchand L, Lum A, Alavanja M, Hayes RB, et al. Collection of genomic DNA from adults in epidemiological studies by buccal cytobrush and mouthwash. Cancer Epidemiol Biomarkers Prev 2001;10:687–96.
15. Packer BR, Yeager M, Staats B, Welch R, Crenshaw A, Kiley M, Eckert A, Beerman M, Miller E, Bergen A, Rothman N, Strausberg R, et al. SNP500Cancer: a public resource for sequence validation and assay development for genetic variation in candidate genes. Nucleic Acids Res 2004;32:D528–32.
16. Wang N, Akey JM, Zhang K, Chakraborty R, Jin L. Distribution of recombination crossovers and the origin of haplotype blocks: the interplay of population history, recombination, and mutation. Am J Hum Genet 2002;71:1227–34.
17. Lan Q, Chapman RS, Schreinemachers DM, Tian L, He X. Household stove improvement and risk of lung cancer in Xuanwei, China. J Natl Cancer Inst 2002;94:826–35.
18. DeMarini DM, Landi S, Tian D, Hanley NM, Li X, Hu F, Roop BC, Mass MJ, Keohavong P, Gao W, Olivier M, Hainaut P, et al. Lung tumor *KRAS* and *TP53* mutations in nonsmokers reflect exposure to PAH-rich coal combustion emissions. Cancer Res 2001;61:6679–81.
19. Schaeffer L, Moncollin V, Roy R, Staub A, Mezzina M, Sarasin A, Weeda G, Hoeijmakers JH, Egly JM. The ERCC2/DNA repair protein is associated with the class II BTF2/TFIIH transcription factor. EMBO J 1994;13:2388–92.
20. Coin F, Bergmann E, Tremeau-Bravard A, Egly JM. Mutations in *XPB* and *XPD* helicases found in xeroderma pigmentosum patients impair the transcription function of TFIIH. EMBO J 1999;18:1357–66.
21. Hou SM, Falt S, Angelini S, Yang K, Nyberg F, Lambert B, Hemminki K. The *XPD* variant alleles are associated with increased aromatic DNA adduct level and lung cancer risk. Carcinogenesis 2002;23:599–603.
22. Lunn RM, Helzlsouer KJ, Parshad R, Umbach DM, Harris EL, Sanford KK, Bell DA. *XPD* polymorphisms: effects on DNA repair proficiency. Carcinogenesis 2000;21:551–5.
23. Matullo G, Palli D, Peluso M, Guarrera S, Carturan S, Celentano E, Krogh V, Munnia A, Tumino R, Polidoro S, Piazza A, Vineis P. *XRCC1, XRCC3, XPD* gene polymorphisms, smoking and ^{32}P-DNA adducts in a sample of healthy subjects. Carcinogenesis 2001;22:1437–45.
24. Qiao Y, Spitz MR, Guo Z, Hadeyati M, Grossman L, Kraemer KH, Wei Q. Rapid assessment of repair of ultraviolet DNA damage with a modified host-cell reactivation assay using a luciferase reporter gene and correlation with polymorphisms of DNA repair genes in normal human lymphocytes. Mutat Res 2002;509:165–74.
25. Au WW, Salama SA, Sierra-Torres CH. Functional characterization of polymorphisms in DNA repair genes using cytogenetic challenge assays. Environ Health Perspect 2003;111:1843–50.
26. Palli D, Russo A, Masala G, Saieva C, Guarrera S, Carturan S, Munnia A, Matullo G, Peluso M. DNA adduct levels and DNA repair polymorphisms in traffic-exposed workers and a general population sample. Int J Cancer 2001;94:121–7.
27. Pastorelli R, Cerri A, Mezzetti M, Consonni E, Airoldi L. Effect of DNA repair gene polymorphisms on BPDE-DNA adducts in human lymphocytes. Int J Cancer 2002;100:9–13.
28. Duell EJ, Wiencke JK, Cheng TJ, Varkonyi A, Zuo ZF, Ashok TD, Mark EJ, Wain JC, Christiani DC, Kelsey KT. Polymorphisms in the DNA repair genes *XRCC1* and *ERCC2* and biomarkers of DNA damage in human blood mononuclear cells. Carcinogenesis 2000;21:965–71.
29. Vodicka P, Kumar R, Stetina R, Sanyal S, Soucek P, Haufroid V, Dusinska M, Kuricova M, Zamecnikova M, Musak L, Buchancova J, Norppa H, et al. Genetic polymorphisms in DNA repair genes and possible links with DNA repair rates, chromosomal aberrations and single-strand breaks in DNA. Carcinogenesis 2004;25:757–63.
30. Chen S, Tang D, Xue K, Xu L, Ma G, Hsu Y, Cho SS. DNA repair gene *XRCC1* and *XPD* polymorphisms and risk of lung cancer in a Chinese population. Carcinogenesis 2002;23:1321–5.
31. Liang G, Xing D, Miao X, Tan W, Yu C, Lu W, Lin D. Sequence variations in the DNA repair gene *XPD* and risk of lung cancer in a Chinese population. Int J Cancer 2003;105:669–73.
32. King CR, Yu J, Freimuth RR, McLeod HL, Marsh S. Interethnic variability of *ERCC2* polymorphisms. Pharmacogenomics J 2005;54: 54–9.
33. Xing D, Tan W, Wei Q, Lin D. Polymorphisms of the DNA repair gene *XPD* and risk of lung cancer in a Chinese population. Lung Cancer 2002;38:123–9.
34. Cheo DL, Burns DK, Meira LB, Houle JF, Friedberg EC. Mutational inactivation of the xeroderma pigmentosum group C gene confers predisposition to 2-acetylaminofluorene-induced liver and lung cancer and to spontaneous testicular cancer in Trp53$^{-/-}$ mice. Cancer Res 1999;59:771–5.
35. Khan SG, Metter EJ, Tarone RE, Bohr VA, Grossman L, Hedayati M, Bale SJ, Emmert S, Kraemer KH. A new xeroderma pigmentosum group C poly(AT) insertion/deletion polymorphism. Carcinogenesis 2000;21:1821–5.
36. Khan SG, Muniz-Medina V, Shahlavi T, Baker CC, Inui H, Ueda T, Emmert S, Schneider TD, Kraemer KH. The human *XPC* DNA repair gene: arrangement, splice site information content and influence of a single nucleotide polymorphism in a splice acceptor site on alternative splicing and function. Nucleic Acids Res 2002;30:3624–31.
37. Wang YG, Xing DY, Tan W, Wang LJ, Tang PZ, Lin DX. Poly(AT) polymorphism in DNA repair gene *XPC* and lung cancer risk. Zhonghua Zhong Liu Za Zhi 2003;25:555–7.
38. Marin MS, Lopez-Cima MF, Garcia-Castro L, Pascual T, Marron MG, Tardon A. Poly(AT) polymorphism in intron 11 of the *XPC* DNA repair gene enhances the risk of lung cancer. Cancer Epidemiol Biomarkers Prev 2004;13:1788–93.
39. Wacholder S, Chanock S, Garcia-Closas M, El Ghormli L, Rothman N. Assessing the probability that a positive report is false: an approach for molecular epidemiology studies. J Natl Cancer Inst 2004;96:434–42.